Li, Steven and Yang, Qianqian (2009) ***The relationship between implied and realized volatility : evidence from the Australian stock index option market.*** Review of Quantitative Finance and Accounting, 32(4). pp. 405-419.

The Relationship between Implied and Realized Volatility: Evidence from the Australian Stock Index Option Market

Steven Li* and Qianqian Yang^

*International Graduate School of Business, University of South Australia, Adelaide, SA 5001, Australia

^School of Economics and Finance, Queensland University of Technology, Brisbane, QLD4001, Australia

Abstract

This paper examines the relationship between the volatility implied in option prices and the subsequently realized volatility by using the S&P/ASX 200 index options (XJO) traded on the Australian stock exchange (ASX) during a period of five years. Unlike the stock index options such as the S&P 100 index options in the US market, the S&P/ASX 200 index options are traded infrequently, in low volumes, and with long maturity cycle. This implies that the error-in-variables problem for measurement of implied volatility is more likely to exist. After accounting for this problem by instrumental variable method, it is found that both call and put options implied volatilities are superior to historical volatility in forecasting future realized volatility. Moreover, implied call volatility is nearly an unbiased forecast of future volatility.

Keywords: Index options, implied volatility, realized volatility

* Corresponding author. Email: Steven.Li@unisa.edu.au, Tel: +61 8 830 29071.
^ Email: q.yang@qut.edu.au, Tel: +61 7 31381658.

1. Introduction

An implied volatility is the volatility implied by an option price observed in the market using on an option pricing model. Assuming that the market is efficient and the option pricing model is valid, then implied volatility should be an unbiased and efficient prediction of future realized volatility over the remaining life of the option. That is, implied volatility should subsume all the information content contained in all other variables used to explain future realized volatility. Hence, the relation between implied and realized volatility is a joint test of market efficiency and applicability of option pricing model.

The relationship has been an important research topic and many studies have been devoted to this topic. The first study is done by Latane and Rendleman (1976). They use closing prices of options and stocks for 24 companies whose options are traded on the Chicago Board Options Exchange (CBOE) and conclude that implied volatility outperforms historical volatility in forecasting future realized volatility. After that, Chiras and Manaster (1978) and Beckers (1981) also obtain the same conclusion based on a broader sample of CBOE stock options. It should be noted that these studies concentrate on static cross-sectional rather than time-series forecasts.

With the availability of sufficient time series data, later studies start to focus on testing the relation of implied volatility and realized volatility in a dynamic setting. Day and Lewis (1992) examine options on S&P 100 index (OEX options) between 1983 and 1989, and report that implied volatility does not contain more information content on future realized volatility than historical volatility. However, this study ignores the term structure of volatility in the measurement of realized volatility, which is not matched with the remaining life of options. Canina and Figlewski (1993) conduct the regression of realized volatility over the remaining life of option on the implied volatility of S&P 100 index options from 1983 to 1986. Surprisingly, they find that “implied volatility has virtually no correlation with future return volatility”. Lamoureux and Lastrapes (1993) examine options on individual stocks from 1982 to 1984, and find that information contained in historical volatility about future realized volatility is more than that contained in implied volatility. This result is also consistent with that of Day and Lewis (1992).

Responding to the mixed conclusions in the previous studies on individual stock options and index options, Jorion (1995) points out that there are two possible explanations: one is that the test procedure is faulty; the other is that the option markets are inefficient. In contrast with individual stock options and index options, he uses options on foreign currency futures traded on Chicago Mercantile Exchange (CME), and concludes that implied volatility is an efficient but biased forecast of future realized volatility. Further, Fleming (1998) examines the S&P 100 index options from 1985 to 1992, and indicates that the implied volatility is an upward biased forecast, but also that it contains more information regarding to future realized volatility than historical volatility.

All of these studies above are based on overlapping datasets and suffer from the serial correlation problem. That is, historical volatility may contain part of information in the future realized volatility, which leads to overstating the explanatory power of historical volatility.

To address this problem, Christensen and Prabhala (1998) introduce a new sampling procedure which uses non-overlapping volatility series. That is, exactly one implied-volatility is responding to one realized-volatility for each time period under consideration. With this sampling procedure and a longer volatility series from 1983 to 1995, they find that implied volatility of S&P 100 index options outperforms historical volatility in predicting future realized volatility. These conclusions are further enhanced by Christensen and Hansen (2002) by using trade weighed averages of implied volatilities for both OEX puts and calls.

Several studies on other index options have been carried out using the same sampling procedure as Christensen and Prabhala (1998). Hansen (2001) analyses the information content of options on the Danish KFX share index. This option market is very illiquid compared to the OEX options market. It is shown that when error-in-variable problem is controlled by instrumental variable techniques, call implied volatility still contains more information about future realized volatility than historical volatility in such an illiquid option market. More recently, Shu and Zhang (2003) examine the options on S&P 500 index, and also report that implied volatility outperforms the subsequently historical index return volatility. Szakmary *et al* (2003) examine 35 futures options markets across eight separate exchanges and find that for a large majority of the commodities studied, implied volatility is a better predictor of future realized volatility than historical volatility.

In sum, most existing studies have focused on options in the US market and have tended to conclude that implied volatility outperforms historical volatility as a predictor of the subsequently realized volatility over the remaining life of an option. To the authors' best knowledge, no such investigation has been carried out for the S&P/ASX 200 index options market on the Australian Stock Exchange (ASX; now called Australian Securities Exchange). This paper aims to fill this gap in the literature.

We follow the Christensen and Prabhala (1998) approach to investigate the relation between implied and realized volatility by using the ASX stock index options. Our results indicate that implied volatility of ASX index options is an unbiased predictor of future realized volatility. Furthermore, implied volatility has similar predictive power as historical volatility in forecasting future realized volatility.

The rest of the paper is organized as follows. The next section describes the data and sampling procedure. Section 3 illuminates the methodology and provides descriptive statistics for these series. Section 4 presents the empirical results, and Section 5 concludes the paper.

2. Data and sampling procedure

2.1 Data description

This paper is concerned with S&P/ASX 200 index options (XJO) traded in the Australian Stock Exchange (ASX)[1]. The S&P/ASX 200 index options are European and with quarterly expiry cycle: March, June, September and December. The expiration day is the third Thursday[2] of the expiry month or the following business day when an expiry Thursday happens to be a public holiday. On an expiry date, trading of the expiring contracts ceases at 12 noon. This means trading continues after the settlement price has been determined. They are cash settled based on the opening prices of the stocks in the underlying index on the morning of the last trading date. The options are quoted in index point and each index point carries a multiplier of A$10.

A number of changes in the index options market on the ASX are relevant to this analysis. In November 15, 1985, the ASX first listed options on All Ordinaries Index (XAO), which was the main benchmark for stocks listed on the ASX. On April 3,

[1] There are other index options traded on ASX, namely: S&P/ASX 200 property trust index options (XPJ) and S&P/ASX 50 index options (XFL) (see www.asx.com.au). The S&P/ASX 200 index options are chosen for this study because they are the most popular and liquid index options on the ASX

[2] It was the third Friday before September 2004.

2000, a new index, namely S&P/ASX 200 index, was introduced by and became the new benchmark for the Australian stock market. Consequently, the underlying index for the ASX index options also switched from All Ordinaries Index to S&P/ASX 200 index on April 3, 2000. However, during the period from April 3, 2000 and March 31, 2001, a continuation of the former All Ordinaries Index was calculated and disseminated by the ASX to allow for the maturity of futures contracts based on the superseded All Ordinaries Index. During this period the ASX listed index options on the All Ordinaries Index. Thus, from March 31, 2001, S&P/ASX 200 index has been formally used as the underlying asset of index options on the ASX. Additionally, for the reason of thin trading, All Ordinaries index options were delisted twice[3] and finally relaunched until November 8, 1999.

Given these complicated historical changes and our research purpose, we use the data of S&P/ASX 200 index options over a five-year period from April 2, 2001 to March 16, 2006. The period before April 2, 2001 is omitted to avoid possible excessive market movements due to the change in the underlying index and the infrequent trading in an emerging index option market. The daily index options data are provided by ASX[4], consisting of trading date, expiration date, close prices, strike prices and trading volume for each trading options.

The corresponding daily close index levels are obtained from Bloomberg. For the risk-free interest rate r_t, we use Australian 90-day Bank Accepted Bill (BAB) rate as a proxy, since this is probably close to the rate faced by option traders and the maturity of the BAB matches the S&P/ASX option's maturity well. The interest rate data are obtained from Reserve Bank of Australia (RBA). For the purpose of estimating

[3] The first time it was delisted on May 26, 1988 and relaunched on December 8, 1993. The second time it was delisted on September 29, 1996 and relaunched until November 8, 1999.

[4] We are grateful to ASX for providing us with the data.

dividend during the life of options, daily closing prices of SFE SPI 200 futures are also collected from Bloomberg.

2.2 Sampling procedure

To obtain a relatively accurate measurement of implied volatility, the options must be chosen carefully. Our sampling procedure differs from previous studies in that we also take into account the trading volume of the options, and hence our sampling procedure is more suitable to illiquid markets. The following sampling criteria are applied:

(1) Options must be traded on a business day close to but after an expiry date, and have expiration on the next expiry date;

(2) Options must be close to be at-the-money (ATM), i.e. $S_t / X_t \in (0.95, 1.05)$, where S_t is the index level and X_t is the exercise price of option;

(3) Options must be traded actively, i.e., they must have a relatively high trading volume.

Criterion (1) is used to avoid the overlapping of data[5]. The S&P/ASX 200 index options expire on the third Thursday of the settlement month: March, June, September and December, hence four non-overlapped observations can be obtained each year for call or put options. Criterion (2) is used as the option pricing model for calculating implied volatility is more accurate for close to be ATM options[6]. Thus implied volatilities derived from these options are less likely to have measurement errors.

[5] Christensen and Prabhala (1998), Hansen (2001), Christensen and Hansen (2002), and Shu and Zhang (2003) also use non-overlapping time series.
[6] For example, the volatility smile has been well documented in the literature, see e.g. Hull (2003).

However, some close to ATM options may be thinly traded and their prices do not necessarily reflect the market prices, thus Criterion (3) is needed.

Following the criteria described above, the sampling procedure is as follows. Let t be the business day that immediately follows an expiry date. On day t, the closing prices (C_t and P_t) and strike prices($X_{t,c}$ and $X_{t,p}$) are recorded for a call option and a put option, each of which expires on the next expiry date $t+1$ and has highest trading volume among the close to ATM options. The corresponding underlying index level (S_t) is also recorded. To estimate the dividend paid during the life of an option, the closing price of SFE SPI 200 futures (F_t), which has the same time to maturity as the recorded option, is also recorded on day t. The next call option is sampled on the business day that immediately follows the next expiry date. This sampling date is labelled $t+1$. Similarly we record values of $C_{t+1}, P_{t+1}, X_{t+1,c}, X_{t+1,p}, S_{t+1}, F_{t+1}$. The whole sequence of call and put options contracts is constructed in this manner.

Due to the fact that the trades in ASX options markets are sometimes not active, there are a few complications in the data collection. In some instances, no options with three months to maturity are traded on the business day that immediately follows the expiry date, or the options traded on that day deviate too much from being at-the-money. In these cases, we sample the options contracts which are close to at-the-money on the nearest following business day. In some other cases, we may not be able to find the suitable call and put options on the same day. Thus we move to the next closest business day to satisfy all the sampling requirements for both call and put options.

Applying the above sample collection procedure, we end up with 20 observations for call and put options, respectively.

3. Methodology

3.1 Variable definitions and some technical issues

- Time to maturity

In practice, the time for paying interest is based on calendar days, while the time for the life of an option is based on trading days. French (1984) suggests that, when calculating the option's price, time for paying interest and time for option's life should be measured separately. Hull (2003) suggests that, however, in practice, these two measurements do not have much difference except for options with very short life. The option's life in our study is about three months. Thus we use the following definition.

Time to maturity, T_t, is measured by the number of the trading days between the day of trade and the day immediately prior to expiry day divided by the number of trading days per year which is taken as 252. Note that we do not take the expiry day from the trading days into account because all expiring contracts cease at 12:00 noon on the expiry day and the cash settlement price is calculated on the expiry morning.

- Implied volatility

Implied volatility is the volatility implied by an option price observed in the market based on an option pricing model. Thus our study is dependent on the option pricing model which we use. Given that ASX index options are European options, it is natural for us to use Black-Scholes-Merton (BSM) model (Black and Scholes, 1973 and Merton 1973). However, the dividends during the remaining life of the options must be accounted for.

Let $\sigma_{i,t}$ denote the volatility implied by an index options prices and q_t denote the annualized continuous dividend yield during the remaining life of the option, then the BSM formulae for call and put can be expressed as:

$$C_t = S_t e^{-q_t T_t} N(d_1) - X_t e^{-r_t T_t} N(d_2) \tag{1}$$

and

$$P_t = X_t e^{-r_t T_t} N(-d_2) - S_t e^{-q_t T_t} N(-d_1) \tag{2}$$

respectively, where

$$d_1 = (\ln(S_t / X_t) + (r_t - q_t)T_t + \sigma_{i,t}^2 T_t / 2) / \sigma_{i,t}\sqrt{T_t}\ ,$$
$$d_2 = d_1 - \sigma_{i,t}\sqrt{T_t}\,.$$

Hence, we need an estimate of the dividend yield q_t. To this end, we use the SFE index futures contracts which have exactly the same expiry circle, expiry date and underlying asset with the XJO options. The cost of carry model (see, e.g. Hull, 2003) gives

$$F_t = S_t e^{(r_t - q_t)T_t} \tag{3}$$

Combining (3) with (1) and (2) yields

$$C_t = F_t e^{-r_t T_t} N(d_1) - X_t e^{-r_t T_t} N(d_2) \tag{4}$$

and

$$P_t = X_t e^{-r_t T_t} N(-d_2) - F_t e^{-r_t T_t} N(-d_1) \tag{5}$$

where

$$d_1 = (\ln(F_t / X_t) + \sigma_{i,t}^2 T_t / 2) / \sigma_{i,t}\sqrt{T_t}$$
$$d_2 = d_1 - \sigma_{i,t}\sqrt{T_t}$$

By solving (4) and (5) numerically[7], we can obtain the implied volatility series for call and put options.

- *Realized volatility*

The realized volatility can be measured by the sample standard deviation of the daily index returns over the remaining life of an option.

Let n be the number of trading days before the expiration of an option, S_i be the index level, and R_i denote log-return on the ith day during the remaining life of the option. Then we have

$$R_i = \ln(S_i \;/ S_{i-1})$$

for $i = 2$, 3 … n. Thus, the annually realized volatility can be expressed as:

$$\sigma_{r,t} = \sqrt{\frac{252}{n-2}\sum_{i=2}^{n}(R_{i,t} - \overline{R}_t)^2}\ , \tag{6}$$

where $\overline{R}_t$ denote the mean of daily log-returns of the index.

- *Historical volatility*

In the previous studies, historical volatility at time t is often defined as realized volatility at time $t-1$. However, in our sample, the time to maturity ranges from 53 to 63 days. If we follow the measurement above, the information contained in the gap between two consecutive contracts will be ignored. We believe the more recent data contains more relevant information about the future. Thus we use a different definition of historical volatility as follows. For a given option contract with T days to

[7] We use an iteration procedure to calculate implied volatily. Explicit formulae can also be employed, see e.g. Li (2005).

maturity at time *t*, the corresponding historical volatility is calculated by using the daily returns of the period going back *T* days from time *t*. That is,

$$\sigma_{h,t-1} = \sqrt{\frac{252}{T-2}\sum_{i=2}^{T}(R_{i,t-1} - \overline{R}_{t-1})^2}\ , \tag{7}$$

where $\overline{R}_{t-1}$ denote the mean of daily index log-returns during the period t-1.

3.2 Sample descriptive statistics

Table 1 presents descriptive statistics for the level and natural logarithm series for implied-, realized- and historical- volatility.

<<Insert Table 1 about here>>

Table 1 clearly indicates that the average (log) implied volatility is larger than the average (log) realized and average (log) historical volatility for both call and put options series. This indicates that the BSM model tends to overprice both close to ATM call and put options on average, relative to historical volatility level. A similar finding has been demonstrated for options on S&P 100 index (Christensen and Prabhala, 1998), for options on S&P 500 index (Lin *et al*., 1998) and for options on Danish KFX index (Hansen, 2001).

The average (log) historical volatility is sightly higher than the average (log) realized volatility. This may be due to the excessive volatility on the expiration days. This may be due to the fact that in the measurement of realized volatility we exclude the expiration day while in the measurement of historical volatility we include it.

Comparing the call series and put series, we notice that the average (log) implied volatility for put option series is sightly higher than that for call option series. This may be due to the fact, as suggested by Harvey and Whaley (1991; 1992), that longing index put option is a convenient and relative cheap method for hedging. Therefore, buying pressure on index put options is larger than that on index call options. Consequently, implied put volatility is higher than implied call volatility on average.

Turning to the standard deviations, we find that realized volatility is always more volatile than implied call and put volatility in both level and log series, which accords with the notion that implied volatility is a smoothed expectation of future realized volatility.

Now let us consider the distributions of the level series in contrast to the log series for all three volatility measures. Table 1 reveals that the skewness of each log-volatility series is closer to zero than that of the corresponding level volatility series, while the kurtosis of each log-volatility series departs more from three than that of level volatility series. Overall, according to the Jarque-Bera test, it appears that the log-volatility series conform better to the normal distribution than the level volatility series. For this reason, we will focus on the log-volatility series in the following sections. Throughout the paper, we denote by rv_t, iv_t ($iv_{t,c}$, $iv_{t,p}$) and hv_{t-1} the natural logarithm of the realized volatility, implied (call and put) volatility, and historical volatility, respectively.

3.3 Measurement errors in implied volatility

Before we assess the empirical results, we should be aware that there are several sources of measurement errors which may afflict the estimation of implied volatility.

First, option prices, closing index levels and futures prices may be non-synchronous. This is either because the closing times for the three markets are different, or because the index options are not traded frequently. The closing option prices may correspond to a trade taking place before the market is closed. However, the index levels and futures contracts do not suffer from such illiquid problems. For examples, when some good news enters into the market between the trade of option and the time when closing index level and futures price are recorded, then the recorded index level or futures price will be higher than the index level or futures price simultaneously corresponding to the option price, indicating that implied call (put) volatility underestimates (overestimates) the true implied volatility. A similar situation can happen with bad news which may lead to deviations in the opposite direction. In reality, good news and bad news come randomly, and hence the two effects can offset each other and the computed implied volatility will not deviate consistently from the true volatility. However, this non-synchronous measurement does cause an *errors-in-variables* problem (EIV), which leads to the correlation between the explanatory variable and the error term in our subsequent regressions.

Second, the BSM option pricing model assumes that the index level follows a log-normal distribution with constant volatility during the life of the option. In the real-world market, this assumption could be violated, for instance, due to the jumps in the index level. Hence, BSM model can be misspecified and implied volatility computed from BSM model can be consequently misspecified. Shu and Zhang (2003) compare implied volatility computed from BSM model with the one implied from Heston (1993) stochastic volatility model, and conclude that implied volatility computed from BS model still has higher explanatory power than that computed from Heston model. Thus, up to now, BSM model may still be the best model for estimating volatility implied in the option prices.

Third, it should be noted that the dividend adjustments are required for applying the original BSM option pricing model. To this end, we use the relation between futures prices and spot prices, namely cost of carry model. However, in practice, the cost of carry model may not hold and thus it may contribute to the EIV problem.

Finally, the XJO options have a three month maturity cycle which implies that we have to assume that the index volatility is a constant over the three month period. However, in practice, a great deal of empirical studies has shown that volatility is not constant and follows its own stochastic process. Hence, the longer maturity period may exacerbate the EIV problem comparing with the S&P 100 index options in the US market whose maturity cycle is only 1 month.

In sum, it is more likely that the EIV problem exists with the XJO options series and thus it must be accounted for when we assess the relation between implied volatility and realized volatility.

4. Empirical results

Two types of estimation methods are conducted to assess the relation between realized and implied volatility. One is the conventional analysis, namely Ordinary Least Squares (OLS) method. To control for the EIV problem, the *Instrumental Variables* (IV) method is employed.

4.1 Conventional method - OLS estimates

To explore the information content of implied volatility by OLS estimates, we run the following regressions for:

$$rv_t = \alpha_0 + \alpha_i iv_t + \varepsilon_t \,, \tag{8}$$

$$rv_t = \alpha_0 + \alpha_h hv_{t-1} + \varepsilon_t \,, \tag{9}$$

$$rv_t = \alpha_0 + \alpha_i iv_t + \alpha_h hv_{t-1} + \varepsilon_t \,, \tag{10}$$

where rv_t, $iv_t (= iv_{t,c}$ or $iv_{t,p})$ and hv_{t-1}, respectively, denote the natural logarithm of the realized volatility, implied volatility and historical volatility.

There are a few testable hypotheses of main interest (see e.g., Christensen and Hansen, 2002). Firstly, if implied volatility contains *some* information about future realized volatility, then the coefficient of the implied volatility α_i in both (8) and (10) should be nonzero. Secondly, if implied volatility is an *unbiased* forecast of future realized volatility, then the coefficient of implied volatility α_i should be equal to one and the coefficient of intercept α_0 should be equal to zero in both (8) and (10). Finally, if implied volatility is an *informationally efficient* predictor of future realized volatility, i.e., implied volatility efficiently impounds all information to predict future volatility, then the coefficient of historical volatility α_h in (10) should be 0, and the error term ε_t should be white noise and thereby uncorrelated with any explanatory variable in the market's information set.

<< Insert Table 2 about here>>

Table 2 presents the results of OLS estimates of regressions (8)-(10). Before we look at the significance of individual coefficient estimates, we first use F-test to check the overall significance of each regression. According to the F-statistics, we find that regression (8) is significant at 5% level for put series, but not significant at this level for call series. This result appears to indicate that implied put volatility contains some information about future realized volatility, while implied call volatility has no significant relation with future realized volatility and thus can not be used as a market forecast of future index volatility. If this result were true, the joint hypothesis of option market efficiency and applicability of BSM model would be rejected. However, as noted previously, this result might be biased because of the EIV problem.

Regression (9) is significant at 5% level, indicating that historical volatility can be used to forecast future volatility. Regressions (10) are not significant at 5% level. This may be due to the EIV problem or the high correlation between the independent variables in the regressions, namely the multicollinearity problem. The correlation matrix in Table 3 provides some evidence on the presence of the multicollinearity problem. The correlation coefficient is 0.70 between implied call and historical volatility, and 0.76 between implied put and historical volatility.

<<Insert Table 3 about here>>

Now let us turn to the significance of individual coefficient estimates. In regression (8) and (9), both coefficients of the historical volatility and implied put volatility are significant at 5% level. The coefficient of historical volatility is around 0.48 and that of implied put volatility is 0.66, indicating that both of them are biased predictor of the future realized volatility. The *Adjusted-R^2* of regression (8) suggests that implied put volatility explains about 19% of the variation of future volatility, while that for historical volatility in regression (9) is about 18%. It appears that implied put volatility dose not significantly overwhelm historical volatility and only has slightly higher explanatory power than historical volatility in forecasting future volatility.

However, neither the coefficient of implied volatility in (8) or (10) is significant. There are two possible explanations of the fact that historical volatility contains more information about subsequent realized volatility than implied call volatility. One is that historical volatility is indeed more informative than implied call volatility for ASX index options. Since ASX index option market is an illiquid market, volatility implied in the option prices is stale and might not be an unbiased and efficient forecast of future realized volatility. If this explanation were true, then it would reject the joint hypothesis that BS formula is valid and ASX index option market is efficient.

The other explanation is that the OLS results are seriously affected by the measurement errors in the implied volatility, namely, EIV problem. As suggested by Christensen and Prabhala (1998), the presence of EIV problem can lead to a few consequences: the estimated coefficient of implied volatility $\hat{\alpha}_i$ is downward biased, even to zero, and the estimated coefficient of historical volatility $\hat{\alpha}_h$ is upward biased. This suggests that the OLS estimates of regressions (8) and (10) appear to be biased and inconsistent, and hence lead to the incorrect conclusions that XJO implied volatility is biased and inefficient.

Up to this point, our results reveal that XJO implied volatility is neither unbiased nor efficient. Implied put volatility has slightly more predictive power than historical volatility, whether assessed by the *Adjusted-*R^2 of each regression or by the magnitude of the regression slope coefficients. In this respect, XJO index options do not appear to be dramatically different from KFX options (Hansen, 2001), OEX options (Christensen and Hansen, 2002), except that implied call volatility does not have significant relation with future index volatility.

Given these OLS regression results, we therefore need to investigate if EIV problem is significant for the XJO option series, and if so, to take the EIV problem into account in our analysis.

4.2 Alternative method - IV estimates

In this section, we first use a formal test to verify the presence of EIV problem, then we will show whether the previous results can be improved after accounting for the EIV problem by IV method.

4.2.1 Presence of EIV problems

As noted previously, implied volatility iv_t in the regression (8) and (10) might be correlated with error term ε_t and leads to the EIV problem. If EIV problem exists in the OLS regression (8) and (10), then the OLS estimates can be not only biased but also inconsistent. That is, the estimates do not converge to the true population value as the sample size increases infinitely. Thus, the OLS estimates of regression (8) and (10) may yield misleading results. In this case, the remedy is the alternative method, IV method. However, the IV estimates are less efficient than OLS estimates if EIV problem does not exist. Thus we need to test for the presence of EIV problem using the Hausman (1978) test.

The basic idea of Hausman (1978) test is to construct a χ^2 test statistic based on the difference between OLS estimator and IV estimator. However, Davidson and MacKinnon (1989, 1993) suggest that one never need to construct the difference between two estimators to compute the statistic. They propose a simple version to illustrate Hausman (1978) test by useing an auxiliary regression. To carry out the Hausman test, two OLS regressions are run. In the first stage, we regress the suspect variable iv_t on all exogenous variables and instrument variables. In our case, the regression is given as

$$iv_t = \beta_0 + \beta_1 iv_{t-1} + \beta_2 hv_{t-1} + e_t \,. \tag{11}$$

Then in the second stage, we re-estimate regressions (8) and (10) by including the residuals from the first regression as an additional regressor. That is,

$$rv_t = \alpha_0 + \alpha_i iv_t + \alpha_h hv_{t-1} + \alpha_e e_t + \varepsilon_t \tag{12}$$

If the OLS estimates are consistent, then the coefficient on the first stage residuals α_e should not be significantly different from zero.

For call option series, the estimate of α_e is -0.86 and significant at 15% level. For put option series, the coefficient estimate of α_e is -0.40 but not statistically significant. Hence, the presence of EIV problem is confirmed for call option series but not for put option series. Thus the IV method will be applied to call option series in the following analysis.

4.2.2 IV estimates

Now we will show how to account for the EIV problem by IV method. As suggested by Greene (2000), the idea behind IV method is to find out a set of variables, termed instruments, which are highly correlated with the suspect explanatory variables, but uncorrelated with the error term. These instruments are used to eliminate the correlation between the explanatory variables and the error term. The number of instruments must satisfy the order condition for identification, i.e., there must be at least as many instruments as there are coefficients in the regressions. The natural candidates of instruments are lagged implied volatility iv_{t-1} and historical volatility hv_{t-1}, since iv_{t-1} and hv_{t-1} are highly correlated with implied volatility at time *t*, as shown in Table 3, but are quite plausibly uncorrelated with ε_t, the error term associated with implied volatility sampled three months later.

The IV method can be achieved by the two stage least squares procedure (2SLS). In the first stage, we regress implied volatility iv_t on the instrumental variables iv_{t-1} and hv_{t-1} by OLS method:

$$iv_t = \beta_0 + \beta_1 iv_{t-1} + \beta_2 hv_{t-1} + e_t \ . \tag{13}$$

In the second stage, we re-examine the regressions (8) and (10) by replacing implied volatility iv_t with the fitted implied volatility $i\hat{v}_t$ from the first stage regression. That is,

$$rv_t = \alpha_0 + \alpha_i \hat{iv}_t + \varepsilon_t \quad (14)$$

$$rv_t = \alpha_0 + \alpha_i \hat{iv}_t + \alpha_h hv_{t-1} + \varepsilon_t \quad (15)$$

Table 4 reports the results of 2SLS estimates for call option series. Panel A presents the OLS estimates of the first stage regression (13). It appears that constant term, lagged implied call volatility iv_{t-1} and historical volatility hv_{t-1} altogether explains about 44 percent of the total variation of implied call volatility iv_t. The high level of adjusted R^2 makes the second stage IV regression more significant. The coefficient of historical volatility hv_{t-1} is highly significant at 1% level, whilst the constant term and coefficient of lagged implied volatility iv_{t-1} is not significant in the regression (13). Furthermore, these results reveal that implied volatility at time *t* can be predicted at large by historical volatility, which are known to the market in advance of time *t*. Thus, specification (13) provides a way to forecast future implied volatility, and hence future option prices, using past information available in the market.

<<Insert Table 4 about here>>

Panel B in Tables 4 presents the second stage IV estimates of the regression (14) and (15). In the case of univariate regression (14), which removes historical volatility and only includes fitted implied call volatility as explanatory variable, it is found that the coefficients, t-values and adjusted R^2 all increase significantly, comparing to the OLS estimates results in the first line of Table 2. This suggests that the measurement of implied call volatility is indeed affected by EIV problems. In particular, the coefficient of implied call volatility, α_i, in IV estimates increases dramatically from 0.45 to 0.89, and is not significantly different from 1 (with *t*-statistics of -0.28), indicating that implied call volatility is an unbiased forecast of future realized volatility. The t-value increases from 1.58 to 2.31, indicating that implied call volatility now is significant at 5% level to explain future realized volatility. The adjusted R^2 in the IV estimates is also improved, increasing dramatically from 7% to 19%.

In addition, comparing to the OLS estimate results for historical volatility in the third line of Table 2, it is found that, after accounting for the EIV problem, implied call volatility iv_t does appear to contain more information about future realized volatility rv_t than historical volatility hv_{t-1}, whether judged by the adjusted R^2 of each regression or by the magnitude of the regression slope coefficients. In sum, both implied call and put volatility are superior to historical volatility in forecasting future index return volatility, and implied call volatility appears to be less biased than implied put volatility and historical volatility.

In the case of multivariate regression (15), which includes both fitted implied call volatility and historical volatility as regressors, it appears that both regressors are not statistically significant. This is probably due to the weak performance of lagged implied call volatility iv_{t-1} as instrument in the first stage regression (13), which leads to that fitted implied call volatility is highly correlated with historical volatility, and in turn leads to the multicollinearity problem in the second stage regression (15). For this reason, the issue of whether implied volatility is an informationally efficient forecast of future index volatility could not be decided, even after solving for the EIV problem by IV method.

5. Conclusion

We have investigated the relationship between implied volatility and realized volatility by using S&P/ASX 200 index options (XJO) data during the 5-year period of April 2001 to March 2006. We followed the sampling approach proposed by Christensen and Prabhala (1998) to obtain non-overlapping quarterly data on XJO options. Due to the quarterly maturity cycle of the XJO options, we could obtain only 20 observations each for the call series and put series.

Using conventional analysis (OLS method), we found that implied call volatility had no relationship with future volatility, but implied put volatility and historical volatility could be used to forecast future volatility. Furthermore, implied put volatility was less biased and had slightly higher predictive power than historical volatility. However, these results could be misleading because of the presence of an errors-in-variable (EIV) problem resulting from several possible factors.

A Hausman (1978) test confirmed the existence of an EIV problem in the measurement of implied call volatility. To account for this problem, an alternative method (IV method) was utilised. Implied call volatility was indeed found to be an unbiased estimator of future volatility and had slightly higher predictive power than historical volatility.

In sum, both implied call and put volatility were better than historical volatility in forecasting realized volatility, whether assessed by the Adjusted-R2 of each regression or by the magnitude of the regression slope coefficients. Moreover, implied call volatility was nearly an unbiased forecast of future volatility. However, the issue of whether implied volatility was informationally efficient could not resolved, either because of the multicollinearity problem between volatility series, or because of the weak performance of lagged implied call volatility as an instrument for implied call volatility.

This paper has focused on the implied volatility from the Black–Scholes model which is the most widely used option pricing model in practise. There are many other option pricing models, such as the deterministic volatility function (DVF) approach by Dumas et al. (1998), the stochastic volatility models of Heston (1993) and Hull and White (1987), and the jump model of Bates (1996). The relationship between implied volatility and realized volatility in the context of other option pricing models is open to exploration. We leave this for possible future research.

References

Beckers, S., 1981. Standard deviation implied in option prices as predictors of future stock price variability. *Journal of Banking and Finance* 5(3), 363-381.

Black, F., Scholes, M., 1973. The valuation of options and corporate liabilities. *Journal of Political Economy*, 81, 637-654.

Canina, L., Figlewski, S., 1993. The informational content of implied volatility. *Review of Financial Studies* 6, 659-681.

Chiras, D.P., Manaster, S., 1978. The information content of prices and a test of market efficiency. *Journal of Financial Economics* 6, 213-234.

Christensen, B.J., Prabhala, N.R., 1998. The relation between implied and realized volatility. *Journal of Financial Economics* 50, 125-150.

Christensen, B.J., Hansen, C.S., 2002. New evidence on the implied-realized volatility relation, *The European Journal of Finance* 8, 187-205.

Davidson, R., MacKinnon, J. G., 1989. Testing for consistency using artificial regressions. *Econometric Theory* 5, 363-84.

Davidson, R., MacKinnon, J. G., 1993. *Estimation and inference in econometrics*. Oxford University Press.

Day, T., Lewis, C., 1992. Stock market volatility and the information content of stock index options. *Journal of Econometrics* 52, 267-287.

Fleming, J., 1998. The quality of market volatility forecasts implied by S&P 100 index option prices. *Journal of Empirical Finance* 5, 317-345.

French, D.W., 1984. The weekend effect on the distribution of stock prices: Implications for option pricing. *Journal of Financial Economics* 13, 547-559.

Greene, W.H., 2000. *Econometric Analysis*. Prentice Hall International Inc., Upper-Saddle River, NJ.

Hansen, C.S., 2001. The relation between implied and realised volatility in the Danish option and equity markets. *Accounting and Finance* 41, 197-228.

Harvey, C.R., Whaley, R., 1991. S&P 100 index option volatility. *Journal of finance* 46, 1551-1561.

Harvey, C.R., Whaley, R., 1992. Market volatility prediction and the efficiency of the S&P 100 index option market. *Journal of Financial Economics* 31, 43-73.

Hausman, J., 1978. Specification tests in econometrics. *Econometrica* 46, 1251-1271.

Heston, S., 1993. A closed form solution for options with stochastic volatility with applications to bond and currency options. *Review of Financial Studies* 6, 327-343.

Hull, J.C., 2003. *Options, Futures, and other Derivatives*. Prentice Hall International Inc., Upper-Saddle River, NJ.

Jorion, P., 1995. Predicting volatility in the foreign exchange market. *Journal of Finance* 50, 507-528.

Lamoureux, C.G., Lastrapes, W., 1993. Forecasting stock return variance: towards understanding stochastic implied volatility. *Review of Financial Studies* 6, 293-326.

Latane, H., Rendleman, R., 1976. Standard deviation of stock price ratios implied in option prices. *Journal of Finance* 31, 369-381.

Li, S., 2005. A new formula for computing implied volatility, *Applied Mathematics and Computation*, 170, 611-625.

Lin, Y., Strong, N., Xu, G., 1998. The encompassing performance of S&P 500 implied volatility forecasts. Paper presented at the European Finance Association Annual meeting, INSEAD.

Merton, R.C., 1973. Theory of rational option pricing, *The Bell Journal of Economics and Management Science* 4, 141-183.

Shu, J., Zhang, J.E., 2003. The relationship between implied volatility and realised volatility of S&P 500 index. Technical Article 4, *WILMOTT magazine,* 83-91.

Szakmary, A., Ors, E., Kim, J.K., Davidson III, W.N., 2003. The predictive power of implied volatility: Evidence from 35 futures markets. *Journal of Banking and Finance* 27, 2151-2175.

Table 1: Descriptive statistics

This table reports descriptive statistics for quarterly level series and natural logarithm series of implied call and put volatility, realized volatility, and historical volatility for the S&P/ASX 200 stock index. Here, implied volatility is computed each quarter using the Black-Schloles-Merton model; realized volatility is the annualized ex-post daily return volatility (sample standard deviation) of the index over the remaining life of the option. Historical volatility is calculated as the annualized sample standard deviation of the daily index return over a period before the sampling time, but of the same time length as that for calculating realized volatility. Statistics are based on 20 non-overlapping quarterly observations for each series, covering a five-year period from April 2, 2001 to March16, 2006.

Statistics	Implied call volatility	Implied put volatility	Realized volatility	Historical volatility	Log implied call volatility	Log implied put volatility	Log realized volatility	Log historical volatility
Mean	0.1166	0.1330	0.0986	0.0992	-2.1763	-2.0415	-2.3614	-2.3543
Std.Dev	0.0288	0.0304	0.0313	0.0313	0.2371	0.2217	0.3043	0.2994
Skewness	0.7968	0.7358	0.7607	0.8858	0.3079	0.2100	0.3243	0.3716
Kurtosis	2.9936	3.1325	2.6507	3.1169	2.5664	2.7260	2.0570	2.2424
Jarque-Bera	2.1164	1.8194	2.0304	2.6269	0.4726	0.2096	1.0916	0.9386

Table 2: Information content of implied volatility: OLS estimates

This table presents the OLS estimates of regressions (8)-(10) for both call and put series:

$$rv_t = \alpha_0 + \alpha_i iv_t + \varepsilon_t \ ,$$

$$rv_t = \alpha_0 + \alpha_h hv_{t-1} + \varepsilon_t \ ,$$

$$rv_t = \alpha_0 + \alpha_i iv_t + \alpha_h hv_{t-1} + \varepsilon_t \ ,$$

where rv_t denotes the natural logarithm of the daily return realized volatility of the index over the remaining life of option; iv_t ($= iv_{t,c}$ for the call series and $iv_{t,p}$ for the put series) denotes the natural logarithm of Black-Scholes-Merton implied volatility for close to be at-the-money options on the underlying index, measured at time t ; hv_{t-1} denotes the natural logarithm of historical index return volatility during the period before time t and with the same time length as that of the future realized volatility. The data consist of 20 non-overlapping quarterly observations for each series, covering a five-year period from April 2001 to March 2006. Note that DW denotes Durbin-Watson statistic and the numbers in parentheses are *t*-statistics. * denote significant at 5%.

Dependent variable: Log realized volatility rv_t						
Independent variables:				Adj.R^2	DW	F-Statistic
Intercept	$iv_{t,c}$	$iv_{t,p}$	hv_{t-1}			
-1.39*	0.45			7%	1.71	2.49
(-2.24)	(1.58)					
-1.01		0.66*		19%	2.07	5.45*
(-1.74)		(2.33)				
-1.23*			0.48*	18%	2.33	5.14*
(-2.46)			(2.27)			
-1.20	0.042		0.46	13%	2.32	2.43
(-1.95)	(0.11)		(1.49)			
-0.95		0.40	0.25	17%	2.31	2.96
(-1.59)		(0.91)	(0.77)			

Table 3: Correlation Matrix

This table presents a correlation matrix for the dependent and independent variables in regressions (8)-(13), where rv_t denotes the natural logarithm of the daily return realized volatility of the index over the remaining life of option; $iv_{t,c}$ and $iv_{t,p}$ denote the natural logarithm of Black-Scholes-Merton implied volatility for at-the-money call and put options on the underlying index, measured at time t; $iv_{t-1,c}$ and $iv_{t-1,p}$ denote the first lagged value of implied call and put volatility, respectively; and hv_{t-1} denotes the natural logarithm of historical index return volatility during the period before time t and with the same time length as that of the future realized volatility. The data consist of 20 non-overlapping quarterly observations for each series, covering a five-year period from April 2001 to March 2006.

	rv_t	$iv_{t,c}$	$iv_{t,p}$	hv_{t-1}	$iv_{t-1,c}$	$iv_{t-1,p}$
rv_t	1.00					
$iv_{t,c}$	0.36	1.00				
$iv_{t,p}$	0.49	0.79	1.00			
hv_{t-1}	0.48	0.70	0.76	1.00		
$iv_{t-1,c}$	0.27	0.31	0.27	0.33	1.00	
$iv_{t-1,p}$	0.30	0.37	0.27	0.44	0.79	1.00

Table 4: Information content of implied call volatility: 2SLS estimates

This table presents the results of 2SLS estimates by using historical volatility hv_{t-1} and lagged implied call volatility iv_{t-1} as the instruments for implied call volatility iv_t.

Panel A reports the OLS estimates of the first stage regression (13) for call option series,

$$iv_t = \beta_0 + \beta_1 iv_{t-1} + \beta_2 hv_{t-1} + e_t$$

where iv_t denotes the natural logarithm of Black-Scholes-Merton implied volatility for at-the-money options on the underlying index, measured at time t, especially, $iv_{t,c}$ denotes implied call volatility; $iv_{t-1,c}$ denotes the first lagged value of implied call volatility; and hv_{t-1} denotes the natural logarithm of historical index return volatility during the period before time t and with the same time length as that of the future realized volatility.

Panel B presents the IV estimates of the second stage regression (14) and (15),

$$rv_t = \alpha_0 + \alpha_i \hat{iv}_t + \varepsilon_t$$

$$rv_t = \alpha_0 + \alpha_i \hat{iv}_t + \alpha_h hv_{t-1} + \varepsilon_t$$

where $\hat{iv}_t$ is the fitted values from the first stage regression (13). Note that DW denotes Durbin-Watson statistic and the numbers in parentheses are *t*-statistics. ** and * denote significant at 1% and 5%, respectively.

Panel A: First stage regression estimates						
Dependent	Intercept	$iv_{t-1,c}$	hv_{t-1}	Adj.R^2	DW	F-Statistic
$iv_{t,c}$	-0.73	0.09	0.53**	44%	2.29	8.09**
	(-1.66)	(0.50)	(3.61)			
Panel B: Second stage regression estimates						
Dependent	Intercept	*fitted*-$iv_{t,c}$	hv_{t-1}	Adj.R^2	DW	F-Statistic
rv_t	-0.42	0.89*		19%	2.00	5.33*
	(-0.49)	(2.31)				
rv_t	0.35	1.79	-0.51	15%	1.93	2.56
	(0.12)	(0.56)	(-0.29)			